Reviewer #2

Abbaszadeh et al. present a model based study over the Upper Colorado River Basin. Using variety of in situ and satellite observations, a coupled model representing surface and subsurface hydrological processes (ParFlow/LIS-Noah-MP) is compared to a stand-alone LIS-Noah-MP implementation which includes much simpler subsurface representation. The main conclusion is that the coupled model enables more spatial detail to be captured in various states, but that statistical fit metrics with observations change little (or even slightly decline).

Overall the study is well written and illustrated, and I am in favour of its eventual publication in HESS. However, there are a few aspects which I think could be improved. For example:

Thank you for the positive feedback and super useful comments and suggestions. These insights have significantly enhanced both the quality and clarity of our work, helping us refine key aspects of the study and improve its overall readability. We appreciate the time and effort the reviewer dedicated to providing detailed feedback, which has been instrumental in strengthening the presentation and impact of our paper.

The abstract could benefit from being slightly more specific

We further revised the abstract and included more details.

"In general, the results show that the coupled ParFlow-LIS/Noah-MP model produces soil moisture simulations comparable to those of the LIS/Noah-MP model across the entire UCRB. The root mean squared error and correlation coefficients are nearly identical between the two models. However, further analysis—when these metrics are averaged over areas with complex topography—revealed that in regions with a high elevation gradient, the ParFlow-LIS/Noah-MP model outperforms the standalone LIS/Noah-MP model in terms of soil moisture simulation."

The assertions made in the Introduction (L94) that most previous large-scale subsurface modelling studies have not accounted for surface processes isn't really the case any more (as demonstrated by the extensive work on this topic by at least one of the co-authors).

Thank you for the suggestion. To address this comment we have decided to remove this sentence from the text.

Since ParFlow has previously coupled with various LSM and atmospheric models, the motivation for undertaking the specific coupling presented in this paper is not very clear. After reading, it seems that this may come down to the possibility to do Data Assimilation, which the LIS framework enables. However, this is not stated, and it is unclear that any DA was actually conducted.

ParFlow has not been previously coupled with any land surface model within LIS. This study, following our previously published work, is the first to explore the robustness of coupling ParFlow with Noah-MP within LIS to simulate land surface and subsurface hydrologic processes. We also would like to mention that in this study we have not done any data assimilation (DA). As you stated, this is one of the capabilities of the LIS framework, and we believe using the developed coupled system within LIS will enable assimilating satellite observations into PF-LIS/Noah-MP, improving its prediction skills while accounting for uncertainties.

To address this comment, we have revised the introduction section to emphasize the main novelty of this paper and its advantages.

"*The main novelty of this work is to demonstrate the capability of the newly coupled ParFlow and LIS/Noah-MP model in simulating land surface and subsurface hydrologic processes. Although LIS/Noah-MP has been widely used in many studies, its ability to model groundwater processes has been limited. In this study, we assess the performance of the ParFlow groundwater hydrology model when coupled with LIS/Noah-MP, focusing on its ability to simulate subsurface hydrologic processes, such as groundwater and soil water content, and their interactions with land surface processes. It is important to note that the primary goal of this paper is not to compare the performance of the ParFlow-LIS/Noah-MP system to LIS/Noah-MP or any other coupled system. Instead, the focus is on how ParFlow is integrated with LIS/Noah-MP and the resulting improvements, not only in simulating soil moisture (as accurately as LIS/Noah-MP) but also in enabling the simulation of groundwater and other subsurface hydrologic processes, such as pressure head—processes that could not be modeled using LIS alone. Unlike LIS/Noah-MP, the ParFlow-LIS/Noah-MP coupling tracks subsurface water movement by solving the three-dimensional Richards equation, providing a more realistic representation of groundwater storage and water table dynamics*."

The manuscript refers to a paper by Fadji et al. (2024) in which the coupled framework that is applied here is presented. However, it is not listed in the reference list? Is it available to reviewers (e.g. as a pre-print). Since the present paper depends on this framework, this seems like a major omission. In this paper, there seems to be some confusion in the description of how the coupling is achieved.

This paper by Maina et al. (2025) was recently published in the Journal of Advances in Modeling Earth Systems. Here is the link to the paper: https://doi.org/10.1029/2024MS004415. We have updated the references in the revised manuscript.

References

Maina, F. Z., Rosen, D., Abbaszadeh, P., Yang, C., Kumar, S. V., Rodell, M., & Maxwell, R. (2025). Integrating the interconnections between groundwater and land surface processes through the coupled NASA Land Information System and ParFlow environment. Journal of Advances in Modeling Earth Systems (JAMES), 17(2). https://doi.org/10.1029/2024MS004415

Whilst I appreciate the supplemental file, I do not see any comment regarding the availability of the model configuration files and outputs. I strongly encourage the authors to make these materials available to the community so that other can build on their work, and to enhance the transparency and reproducibility of such advanced computational work.

We have added the "Data Availability Statement" section to the revised manuscript.

"Data Availability Statement: ParFlow-LIS is included in the Nasa Land Information System (LIS), an open-source software that can be found at Rosen and Dunlap (2024). NUOPC CAP has been integrated in both ParFlow and LIS. The data set and model configuration for LIS/Noah-MP and ParFlow-LIS/Noah-MP models can be found at Maina (2024)"

References:

Rosen, D., & Dunlap, R. (2024). fadjimaina/NASA-Land-Coupler: ParFlow-LIS (Version release_0.1). Zenodo. https://doi.org/10.5281/zenodo. 14058196

Maina, F. (2024). Integrating the interconnections between groundwater and land surface processes through the coupled NASA Land Information System and ParFlow environment [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.10950634

Please see the attached annotated PDF for my detailed comments.

Line 23: In my view, the syntax "that enable....to be studied" would be preferable.

We revised the sentence as you suggested.

“that enable the Earth's land surface and subsurface hydrologic processes to be studied.”

Line 24: Consider changing to: "The integration... to harness their strengths provides an opportunity..."

We revised the sentence as you suggested.

“The integration of ParFlow and LIS/Noah-MP models to harness their strengths provides an opportunity to simulate surface terrestrial water processes and groundwater dynamics together, while enhancing the accuracy and scalability of hydrological modeling.”

Line 25, 26: This statement is quite general and could perhaps be made more specific. How more precisely could (or does) this approach advance the state of the art or compare with what is typically done? For instance, since ParFlow considers coupled surface-subsurface processes, I am wondering at this stage what LIS/Noah-MP will contribute? I guess the key additional element is the atmosphere?

Thank you for the comment. To clarify, the ParFlow model is a subsurface hydrology model (groundwater model) and LIS/Noah-MP is a land surface model.

Line 29: I understood that ParFlow (and certainly ParFlow-CLM) does this already. So, are you essentially proposing/testing coupling to another LSM?

Yes. In this project, we coupled the ParFlow groundwater model with the LIS/Noah-MP land surface model, and the coupled system is basically a subsurface-surface hydrology model (similar to ParFlow-CLM as you mentioned).

Line 30, 31, 32: Good clear statement on the experimental design.

Thank you for the comment.

Line 34, 35: Consider slight language change: ", but also enables accurate simulation of subsurface hydrologic processes."

We revised the sentence as you suggested.

"This analysis confirmed that integrating ParFlow with LIS/Noah-MP not only enhances the capability of LIS/Noah-MP in estimating land surface processes over regions with complex topography but also enables accurate simulation of subsurface hydrologic processes."

Line 43: Insert "either" for greater clarity. Line 44: Is "be returned" better?

We revised the sentence as you suggested.

"For instance, precipitation that falls on the land surface can either infiltrate the soil and become soil moisture or runoff into nearby streams and rivers. Soil moisture can be returned to the atmosphere through."

Line 51: Change to: ", as well as their interactions..." Also, should feedback be plural?

We revised the sentence as you suggested.

"Climate change can impact surface and subsurface hydrologic processes, as well as their interactions and feedbacks to the atmosphere."

Line 55: Moreover? Or Furthermore?

We revised the sentence as you suggested.

"Furthermore, human activities, such as irrigation and water pumping, can alter the natural behavior of surface–subsurface interaction…"

Line 68: Consider making the specific link with irrigation here; groundwater is pumped and used in irrigation for agriculture.

We revised the sentence as you suggested.

"Groundwater pumping is an important source of water for irrigation in agriculture in the UCRB, particularly when and where surface water availability is limited."

Line 69: I would argue that if pumping is excessive or unsustainable, it *does* lead to aquifer depletion. Same comment could apply above when you write "human activities can modify..." Line 71: Not only of agricultural. Isn't the point that the water is also then unavailable for other applications too?

We revised the sentence as you suggested.

"Excessive pumping can lead to the depletion of aquifers, impacting water availability and the long-term sustainability of agricultural practices."

Line 83: I would avoid the term "layers" here. Components? Facets?

We revised the sentence as you suggested.

"...affect different components of the hydrologic system..."

Line 85: Change to: "these connections" (agreement).

We revised the sentence as you suggested.

"...to better understand these connections has been..."

Line 94, 95: Not sure this is completely true anymore. Most of these citations are rather dated. For example, there's the work applying ParFlow across the CONUS and the coupled global model of de Graaf, etc. Subsurface processes have already been accounted for at a large scale. The question is whether the grid resolutions and subsurface (geological) data used in those studies are useful for local-scale water management applications.

Thank you for the suggestion. Totally agree. To address this comment we have decided to remove this sentence from the text.

Line 98: Check this.

Corrected.

Line 99: Please explain here the relationship between LIS and Noah-MP a little more. Right now, it comes out of nowhere. For instance, is the model really Noah-MP, but it is run within the framework of LIS, which enables integration with satellite observations?

Thanks for the suggestion. We have revised this part of the introduction to address this comment.

"The NASA Land Information System (LIS) is a software framework designed to facilitate the integration of land surface models and satellite remote sensing data for improved understanding and prediction of land surface processes. LIS is a modeling framework that offers a variety of model options. One of the key models that can be run within the LIS framework is the Noah-MP (Multi-Parameterization) model, which is a widely used land surface model. The LIS framework enables the coupling of the Noah-MP model with satellite observations and other data sources, providing a more comprehensive view of land-atmosphere interactions. Specifically, LIS serves as a tool for executing Noah-MP simulations, allowing for real-time integration of remote sensing data and enhancing the model's predictive capabilities."

Line 113, 114: Given the previous coupling efforts, what was the motivation for doing this one? For example, is LIS/Noah-MP expected to be stronger than previously coupled LSM?

ParFlow has not been previously coupled with any land surface model within LIS. This study, following our previously published work (Maina et al. 2025), is the first to explore the robustness of coupling ParFlow with Noah-MP within LIS to simulate land surface and subsurface hydrologic processes. We would also like to note that ParFlow is coupled to a different version of CLM than what is in LIS and that this version is incorporated into ParFlow. That is it's not an external, community modeling platform. So in addition to the differences between CLM and NoahMP, there are software differences too. To address this comment, we have revised the introduction section to emphasize the main novelty of this paper and its advantages.

"*The main novelty of this work is to demonstrate the capability of the newly coupled ParFlow and LIS/Noah-MP model in simulating land surface and subsurface hydrologic processes. Although LIS/Noah-MP has been widely used in many studies, its ability to model groundwater processes has been limited. In this study, we assess the performance of the ParFlow groundwater hydrology model when coupled with LIS/Noah-MP, focusing on its ability to simulate subsurface hydrologic processes, such as groundwater and soil water content, and their interactions with land surface processes. It is important to note that the primary goal of this paper is not to compare the performance of the ParFlow-LIS/Noah-MP system to LIS/Noah-MP or any other coupled system. Instead, the focus is on how ParFlow is integrated with LIS/Noah-MP and the resulting improvements, not only in simulating soil moisture (as accurately as LIS/Noah-MP) but also in enabling the simulation of groundwater and other subsurface hydrologic processes, such as pressure head—processes that could not be modeled using LIS alone. Unlike LIS/Noah-MP, the ParFlow-LIS/Noah-MP coupling tracks* *subsurface water movement by solving the three-dimensional Richards equation, providing a more realistic representation of groundwater storage and water table dynamics*."

References:

Maina, F. Z., Rosen, D., Abbaszadeh, P., Yang, C., Kumar, S. V., Rodell, M., & Maxwell, R. (2025). Integrating the interconnections between groundwater and land surface processes through the coupled NASA Land Information System and ParFlow environment. Journal of Advances in Modeling Earth Systems (JAMES), 17(2). https://doi.org/10.1029/2024MS004415

Line 117: Is it available to reviewers? It does not appear in the reference list!

This paper has recently been published. Please see the following reference. Therefore, we have removed that sentence "This paper has been under review at the time of writing this manuscript" from the text.

References:

Maina, F. Z., Rosen, D., Abbaszadeh, P., Yang, C., Kumar, S. V., Rodell, M., & Maxwell, R. (2025). Integrating the interconnections between groundwater and land surface processes through the coupled NASA Land Information System and ParFlow environment. Journal of Advances in Modeling Earth Systems (JAMES), 17(2). https://doi.org/10.1029/2024MS004415

Line 134: Change to: ". Rather, it..."

Done.

Line 136: Change to: "Allows a realistic representation of..., to be obtained."

We revised the sentence as you suggested.

"This inclusive methodology allows a realistic representation of groundwater dynamics to be obtained, shaped by the underlying geology and topography."

Line 138: As a reader, I am wondering why another approach to model surface processes needs to be proposed?

We used the ParFlow subsurface hydrology model, which solves the partial differential equations (PDEs) governing both surface water and subsurface flow.

Line 143: Change to: "Which ensures?"

Done.

Line 147: It will only "replicate" it if the geometry, material properties, and forcing functions are reasonable. I would prefer a different term like "calculates."

We have revised the sentence as you suggested. Thank you.

Line 149: This is not very clear. I think you are referring to the complexity of the representation of subsurface processes, i.e. the equations. But "subsurface configurations(s)" could also be taken to mean the geology encountered in any given application. There are many factors affecting spin-up time (by the way, why a steady-state solution has to be sought initially in the case of subsequent transient runs has not yet been explained). For instance, it depends on the size of the domain (in 3D) and especially how reasonable the initial conditions applied are. Meanwhile, the domain size, grid resolution, geological complexity, and forcing characteristics also affect the size of the computational problem. I would suggest this part is slightly revised accordingly.

We revised the sentence as you suggested.

"It is important to note that groundwater may take a longer time (compared to soil moisture) to reach steady-state due to the complexity of the representation of subsurface processes. This makes it a computationally intensive problem to solve. There are many factors that influence spin-up time, including the size of the domain (in 3D), the resolution of the grid, the geological complexity, and the characteristics of the forcing data. A steady-state solution is typically sought initially in order to ensure that the model starts from a physically realistic and stable state before transitioning to transient runs. Additionally, the plausibility of the initial conditions applied plays a significant role in determining how quickly the model reaches steady-state. All of these factors contribute to the overall computational complexity of the problem."

Line 154: But these key inputs (initial conditions, boundary conditions, and parameters) also apply for any numerical model, including ParFlow. I find it slightly strange how they are presented here as being somehow unique to LIS.

We revised the sentence as you suggested.

“Land surface modeling within LIS relies on three key inputs: (1) initial conditions, describing the land surface's starting state; (2) boundary conditions, encompassing the atmospheric fluxes or 'forcings' (upper boundary condition) and soil fluxes or states (lower boundary condition); and (3) parameters, which represent the soil, vegetation, topography, and other land surface characteristics. These inputs are not unique to LIS and also apply to other numerical models, such as ParFlow, which similarly requires initial conditions, boundary conditions, and parameters to simulate subsurface and surface processes.”

Line 158: Ok, so this addresses the previous comment about the relationship. You could mention earlier that LIS is a kind of model toolbox, with various model options inside?

We have already addressed this comment above.

Line 168: The possibility of performing?

We have revised this sentence as you suggested.

“The DA embedded within LIS provides the possibility of performing probabilistic simulations…”

Line 175: Exfiltration to streams and rivers, right?
Line 176: Could it be useful to also mention how spatial patterns in subsurface properties are represented under this simple approach?

While Noah-MP captures some spatial variability through parameterized subsurface properties, it does not explicitly simulate lateral groundwater flow or fully resolve spatial heterogeneity at fine scales, which can limit accuracy in complex hydrogeological settings. To address these two comments, we revised the text, please see it below:

"This method tracks variations in groundwater storage based on inflow, known as recharge, and outflows, which include capillary rise and exfiltration to streams and rivers (baseflow). It is important to note that this approach does not explicitly consider complex hydraulic properties, such as hydraulic conductivity, which is typically used in soil moisture modeling and groundwater recharge prediction. Additionally, while Noah-MP captures some spatial variability through parameterized subsurface properties, it does not explicitly simulate lateral groundwater flow or fully resolve spatial heterogeneity at fine scales, which can limit accuracy in complex hydrogeological

settings. As a result, spatial patterns in subsurface properties are not explicitly represented in this simplified approach, which may influence the accuracy of groundwater storage estimates."

Line 180: I know it was not done and would therefore be another study, but would an interesting comparison not have been to compare ParFlow-Noah-MP with ParFlow-CLM?

Yes, this will certainly be one of our future studies. Since this paper focuses on the details and capabilities of the newly developed coupled subsurface-surface hydrology model (ParFlow-LIS/Noah-MP), we have not included or compared other models. However, exploring this comparison will be an important direction for our future research.

Line 181: Did you couple *all* LIS models, or only Noah-MP? If you also coupled CLM, then how does this differ from the previous coupling? (It seems that this would come down to the ability to do DA, which LIS provides?)

No. We have only coupled LIS/Noah-MP with ParFlow. ParFlow and CLM have already been coupled and used in several other studies.

Line 183: # Or runs off? But runoff is also a land surface process. Which model is this simulated using?

ParFlow takes care of the runoff estimation. LIS only calculates the net water flux entering the soil.

Line 184, 185: Specify ParFlow subsurface model.

Revised.

"... used as input to feed the ParFlow subsurface model."

Line 186, 187: Could be rephrased for greater clarity, e.g. "...top four soil layers: the coupled soil zone, in which the two systems communicate."

Revised.

"It should be noted that the land surface model (LIS/Noah-MP) and groundwater model (ParFlow) share the top four soil layers: the coupled soil zone, in which the two models exchange fluxes."

Also, "in which the two models exchange fluxes" might be better than "communicate"?

Addressed above. Thank you so much for reading our paper so carefully and providing super useful and constructive comments. I really appreciate it.

Finally, I would suggest referring immediately to Figure 1 here, which is useful.

Done.

"(See Figure 1)"

Line 192, 193: This is confusing. You say that LIS/Noah calculates the soil moisture, but then next you say this is derived from ParFlow. So is the coupling actually that LIS/Noah-MP calculates recharge to ParFlow, and ParFlow passes back soil moisture content?

Thank you for the precise comment. We have revised this sentence.

"By using saturation data generated by ParFlow as one of its outputs and incorporating the soil layer porosity values, the soil moisture content (θ) is estimated."

Line 204: I guess soil or more generally geological facies-specific storage and porosity? (Below a few meters, there is no soil.)

We have revised this sentence as you suggested.

"ParFlow provides estimates of pressure head and soil saturation, which, along with soil- or more generally geological facies-specific storage and porosity, are used to calculate subsurface storage."

Line 208: So lateral subsurface flow is accounted for in this approach, but lateral surface flow is not? Perhaps specify this if necessary, because so far, there has been little mention of how direct surface runoff is handled.

Thank you for the comment. This is also one of the reviewer #1 suggestion and we have added the following append to the revised manuscript to elaborate more on the ParFlow and how it handles direct surface runoff.

"To address this comment, we have added an appendix to the revised manuscript.

*"**Appendix***

The ParFlow model operates in three distinct modes: (1) variably saturated; (2) steady-state saturated; and (3) integrated watershed flows. This adaptability enhances its utility across a range of hydrological scenarios. Here we summarize each mode following the work of Kollet and Maxwell (2006).

Variably Saturated Flow

ParFlow can operate in variably saturated mode through the well-known mixed form of the Richards' equation:

$$S_s S_w(p)\frac{\partial p}{\partial t} + \phi\frac{\partial\left(S_w(p)\right)}{\partial t} = \nabla q + q_s \qquad (1)$$

$$q = -\, k_s k_r(p)\nabla(p - z) \qquad (2)$$

where S_s is the specific storage coefficient [L-1], S_w is the relative saturation [−] as a function of pressure head p, t is time, ϕ is the porosity of the medium [−], q is the specific volumetric (Darcy) flux [LT−1], k_s is the saturated hydraulic conductivity tensor [LT−1], k_r is the relative permeability [−], which is a function of the pressure head p, q_s is the general source or sink term [T−1] (includes wells and surface fluxes, e.g., evaporation and transpiration). z represents depth below the surface [L]. ParFlow has been utilized for numerical simulations, including the modeling of river–aquifer exchange involving both free-surface flow and subsurface flow. It has also demonstrated efficacy in addressing highly heterogeneous problems under variably saturated flow conditions. For the situations where the saturated conditions are predominant, the steady-state saturated mode in ParFlow becomes a valuable tool.

Steady-State Saturated Flow

The fully saturated groundwater flow equation is expressed as follows:

$$\nabla q - q = 0 \qquad (3)$$

$$q = -\, k_s \nabla P \qquad (4)$$

where P represents the 3-D hydraulic head-potential [L]. ParFlow does include a direct solution option for the steady-state saturated flow that is distinct from the transient solver. When studying more sophisticated or complex processes, such as when simulating a fully coupled system is of interest (i.e., surface and subsurface flow), an overland flow boundary condition is employed.

Overland Flow

Surface water systems are interlinked with the subsurface system; this interaction plays a critical role for rivers. However, explicitly representing the connections between the two systems in numerical simulations is a difficult task. In ParFlow, overland flow is implemented as a two-dimensional kinematic wave equation approximation of the shallow water equations. The continuity equation for two-dimensional shallow overland flow is expressed as follows:

$$\frac{\partial \psi_s}{\partial t} = \nabla\left(\upsilon \psi_s\right) + q_s \qquad (5)$$

where υ *is the depth-averaged velocity vector [LT−1] and* ψ_s *is the surface ponding depth [L]. Ignoring the dynamic and diffusion terms results in the momentum equation, which is known as the kinematic wave approximation:*

$$S_{f,i} = S_{o,i} \qquad (6)$$

The $S_{f,i}$ *and* $S_{o,i}$ *represent the friction [−] and bed slopes (gravity forcing term) [−], respectively.* i *indicates the* x *and* y *directions in the following equations. Therefore, Manning's equation can be used to build a flow depth-discharge relationship as follows:*

$$\upsilon_x = \frac{\sqrt{S_{f,x}}}{n} \psi_s^{2/3} \qquad (7)$$

$$\upsilon_y = \frac{\sqrt{S_{f,y}}}{n} \psi_s^{2/3} \qquad (8)$$

where n *is the Manning roughness coefficient [TL−1/3]. The shallow overland flow formulation (Eq. 9) assumes the vertical averaging of flow depth and disregards any vertical change in momentum within the surface water column. To incorporate vertical flow (from the surface to the subsurface or vice versa), a formulation that couples the system of equations through a boundary condition at the land surface becomes essential. We can modify Equation (5) to include an exchange rate with the subsurface,* q_e*:*

$$\frac{\partial \psi_s}{\partial t} = \nabla\left(\upsilon \psi_s\right) + q_s + q_e \qquad (9)$$

In ParFlow, the overland flow equations are directly coupled to the Richards' equation at the top boundary cell under saturated conditions. Conditions of pressure continuity (i.e., equal pressures at the ground surface for the subsurface and surface domains) and flux at the top cell of the boundary between the subsurface and surface systems are assigned. Setting pressure head in Equation (1) equal to the vertically averaged surface pressure, ψ_s*:*

$p = \psi_s = \psi \qquad (10)$

and the flux, q_e, equal to the specified boundary conditions (for example, Neumann-type boundary conditions):

$q_{BC} = -k_s k_r \nabla(\psi - z) \qquad (11)$

and one solves for the flux term in Equation (10), the result becomes:

$q_e = \frac{\partial \|\psi, 0\|}{\partial t} - \nabla \upsilon \|\psi, 0\| - q_s \qquad (12)$

where the $\|\psi, 0\|$ operator is defined as the greater of the quantities, ψ, 0. Putting the equations (10) and (11) together results in the following relationship:

$-k_s k_r \nabla(\psi - z) = \frac{\partial \|\psi, 0\|}{\partial t} - \nabla \upsilon \|\psi, 0\| - q_s \qquad (13)$

As we see here the surface water equations are represented as a boundary condition to the Richards' equation. For more information about the coupled surface and subsurface flow systems in ParFlow, we refer the interested readers to Kollet and Maxwell (2006)."

References:

Kollet, S. J. and Maxwell, R. M.: Integrated surface-groundwater flow modeling: A free-surface overland flow boundary condition in a parallel groundwater flow model, Adv Water Resour, 29, https://doi.org/10.1016/j.advwatres.2005.08.006, 2006."

Line 209: Suggest "bi-directional exchange."

Revised.

"…facilitating the bi-directional exchange between the land surface and subsurface…"

Line 223: Does this mean that between winters, the maximum SCA fluctuates between these values, or does it mean that within winters the SCA never drops lower than 50,000 and always reaches 280,000 (entire basin)? I guess the former, but as currently written, it is quite ambiguous.

Yes, the former is correct, which means between different winters, the maximum snow-covered area (SCA) varies between 50,000 km^2 and 280,000 km^2—not that every winter follows the exact same range. We have revised this sentence for further clarity.

"The maximum snow-covered area within the UCRB varies between 50,000 km^2 and 280,000 km^2 across different winters during the October–April season."

Line 232: Change to: "Including its climatology and geology" (delete etc.).

We have revised this sentence as you suggested.

"For more information about the UCRB, including its climatology and geology, we refer interested readers to Miller et al. (2016)."

Line 239: So are they not used for any calibration/DA? (In general, I also prefer "evaluation" to "validation.")

No. We have revised the sentence as you suggested.

“In this section we describe all those in-situ observations and satellite products that are used for evaluation of model simulations.”

Line 264: So, is this the second satellite product used? Could perhaps be "signposted" a bit more clearly to readers.

Yes. We have revised this sentence for further clarity.

"The second satellite product that we used in this study for evaluation of coupled PF-LIS model simulation is Anomalies of Terrestrial Water Storage (TWS), derived from the Gravity Recovery and Climate Experiment (GRACE; Tapley et al., 2004) and its successor, GRACE Follow-On (GRACE-FO; Landerer et al., 2020).

Line 280: It would be useful to show the river network too on the left, so that readers can understand the orders of the streams gauged, etc.

To address this comment, we have added the following figure to the revised supplementary file. Thank you for the suggestion.

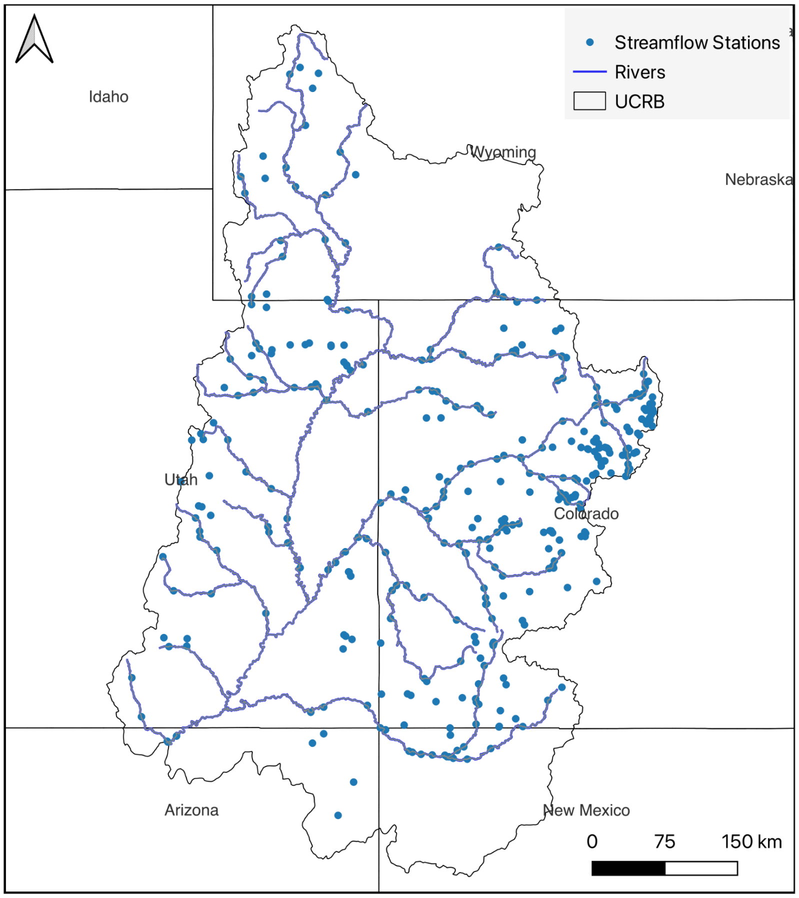

Figure S7. UCRB along with USGS streamflow stations and river networks.

Line 292: Resampled. Using which algorithm?

Using the nearest neighbor method. We have revised that sentence and included this information.

"Land cover information was extracted from the National Land Cover Database (NLCD) at a 30-meter resolution and subsequently resampled using the nearest neighbor method to match the model's 1-kilometer resolution."

Line 301, 302: Please explain more. Which parameters are detailed in which paper? Without this info, there is a risk that the study would not be reproducible.

We have revised the text and included more information.

“The development of the 3D subsurface, which includes soil datasets (e.g., permeability and porosity), unconsolidated, a semi-confining layer, bedrock aquifers, and the 3D model grid, is detailed in Tijerina-Kreuzer et al (2024). The subsurface parameters (e.g. saturated hydraulic conductivity and van Genuchten parameters for the soil and subsurface) are detailed in Yang et al (2023).”

Line 306: What resolution and timestep?

We have revised the text and included this information.

"For the atmospheric forcing data, we use the phase-2 of the North American Land Data Assimilation System (NLDAS-2) product (https://ldas.gsfc.nasa.gov/nldas/v2/forcing). This dataset, available at a spatial resolution of 12.5 km and a temporal resolution of hourly, includes eight variables: precipitation, air temperature, shortwave and longwave radiation, wind speed in two directions (east-west and south-north), atmospheric pressure, and specific humidity."

Line 312: I would rather say "reasonable initial conditions." Line 313: You say above that you had to spin up ParFlow, but here you say you simply took existing output. Please revise for consistency, e.g. "Reasonable initial conditions had to be obtained for both models. To do this, ..."
To address the above two comments, we have revised the text as you suggested.

“Reasonable initial conditions had to be obtained for both models. To do this, The initial condition (i.e., pressure head) for the ParFlow model was directly obtained from Yang et al (2023). who spunup the ParFlow model over the entire CONUS.”

Line 327: Consider adding "..., relative to LIS/Noah-MP alone."

We have revised the text as you suggested.

“...how the coupled system can enhance the modeling of land surface processes and provide a more accurate representation of groundwater storage, relative to LIS/Noah-MP alone.

Line 334: Extending to.

Corrected.

Line 349: I think you mean spatial resolution or specificity.

We have revised the text as you suggested.

“PF-LIS/Noah-MP provides soil moisture data with higher spatial specificity, which can be…”

Line 352: Essentially, it seems that the version with ParFlow is able to "pick out" areas of high soil moisture along river corridors. Can you suggest why this could be? E.g. is it all re-exfiltrating groundwater, or can it be explained by any differences between the representation of topography or river channels in the models?

By integrating surface runoff and subsurface flow processes, ParFlow can simulate the lateral movement of water across the landscape. This dynamic simulation accounts for the redistribution of water due to both surface topography and subsurface properties, leading to a more accurate depiction of soil moisture patterns, particularly in regions influenced by river networks. In addition, the model utilizes high-resolution topographic data to define the land surface and river channels accurately. This precise representation allows ParFlow to identify topographic depressions and convergent zones where water is likely to accumulate, leading to higher soil moisture content. Such detailed modeling ensures that areas prone to saturation, especially along river corridors, are effectively captured.

To further address this comment, we included the above information in the revised text to highlight why the coupled system is able to "pick out" areas of high soil moisture along river corridors.

“The results (shown in Figure 4) also reveal that the coupled system is able to identify the areas of high soil moisture along the river corridors. This is attributed to the ParFlow model. By integrating surface runoff and subsurface flow processes, ParFlow can simulate the lateral movement of water across the landscape. This dynamic simulation accounts for the redistribution of water due to both surface topography and subsurface properties, leading to a more accurate depiction of soil moisture patterns, particularly in regions influenced by river networks. In addition, the model utilizes high-resolution topographic data to define the land surface and river channels accurately. This precise representation allows ParFlow to identify topographic depressions and convergent zones where water is likely to accumulate, leading to higher soil moisture content. Such detailed modeling ensures that areas prone to saturation, especially along river corridors, are effectively captured.”

Line 380, 381: Yes, it is interesting that when aggregated, there is little difference in the fit scores. In one way, this illustrates the limitation, given their limited resolution, of satellite products to evaluate high-resolution simulations.

Yes, this is true. This is an inevitable issue when using satellite products for model evaluation.

Line 399: Elevations.

Corrected.

Line 404: So can LIS/Noah-MP not do this? If so, it would be worth explaining in an earlier section.

Noah-MP can model soil moisture dynamics, but there are limitations in handling complex topography and boundary conditions compared to ParFlow. Noah-MP primarily uses a structured grid system, and while it allows for parameterization of soil moisture processes, it does not inherently solve three-dimensional flow dynamics as ParFlow does. This means that in regions with steep gradients or highly heterogeneous terrain, ParFlow is better suited because it explicitly resolves lateral flow and subsurface heterogeneity. ParFlow's numerical formulation allows for more flexible and realistic boundary condition representations, particularly in irregular terrains. Noah-MP, on the other hand, relies on predefined lookup tables and soil parameterizations that may not fully capture spatial heterogeneity. Noah-MP can incorporate variations in soil moisture, but its hydraulic conductivity representation is more generalized, whereas ParFlow enables explicit simulation of hydraulic gradients and subsurface interactions.

To further address this comment, we have added the above text to the revised manuscript.

"Noah-MP can model soil moisture dynamics, but there are limitations in handling complex topography and boundary conditions compared to ParFlow. Noah-MP primarily uses a structured grid system, and while it allows for parameterization of soil moisture processes, it does not inherently solve three-dimensional flow dynamics as ParFlow does. This means that in regions with steep gradients or highly heterogeneous terrain, ParFlow is better suited because it explicitly resolves lateral flow and subsurface heterogeneity. ParFlow's numerical formulation allows for more flexible and realistic boundary condition representations, particularly in irregular terrains. Noah-MP, on the other hand, relies on predefined lookup tables and soil parameterizations that may not fully capture spatial heterogeneity. Noah-MP can incorporate variations in soil moisture, but its hydraulic conductivity representation is more generalized, whereas ParFlow enables explicit simulation of hydraulic gradients and subsurface interactions."

Line 405, 406: Please compare with what the LIS/Noah-MP does.

Noah-MP does account for both capillary rise and gravitational effects, but its treatment differs from ParFlow in terms of complexity and implementation. Noah-MP incorporates capillary rise through its soil moisture parameterization, which includes the influence of matric potential on soil water movement. However, its representation is more simplified and dependent on predefined soil layers, limiting its ability to dynamically capture variations in soil moisture redistribution, particularly in highly heterogeneous terrains. Noah-MP includes gravitational drainage as part of its hydrology scheme, but it assumes a one-dimensional vertical flow, meaning lateral subsurface flow and complex topographic-driven gravitational effects are not fully resolved. In contrast, ParFlow explicitly solves the three-dimensional Richards' equation, which allows it to capture both vertical and lateral water movement more accurately in complex terrains .

To further address this comment, we have added the above text to the revised manuscript.

"Noah-MP accounts for both capillary rise and gravitational effects, however, its treatment differs from ParFlow in terms of complexity and implementation. Noah-MP incorporates capillary rise through its soil moisture parameterization, which includes the influence of matric potential on soil water movement. However, its representation is more simplified and dependent on predefined soil layers, limiting its ability to dynamically capture variations in soil moisture redistribution, particularly in highly heterogeneous terrains. Noah-MP includes gravitational drainage as part of its hydrology scheme, but it assumes a one-dimensional vertical flow, meaning lateral subsurface flow and complex topographic-driven gravitational effects are not fully resolved. In contrast, ParFlow explicitly solves the three-dimensional Richards' equation, which allows it to capture both vertical and lateral water movement more accurately in complex terrains ."

Line 413: Since you said that the coupled model simulations are limited to the boundaries of the UCRB, I am struggling to interpret this figure. Would it not make sense to remove points outside the catchment? It is also quite difficult to see and compare any differences, apart from via the average statistics.

To address this comment, we have decided to reproduce all these four figures. Thank you for the suggestion.

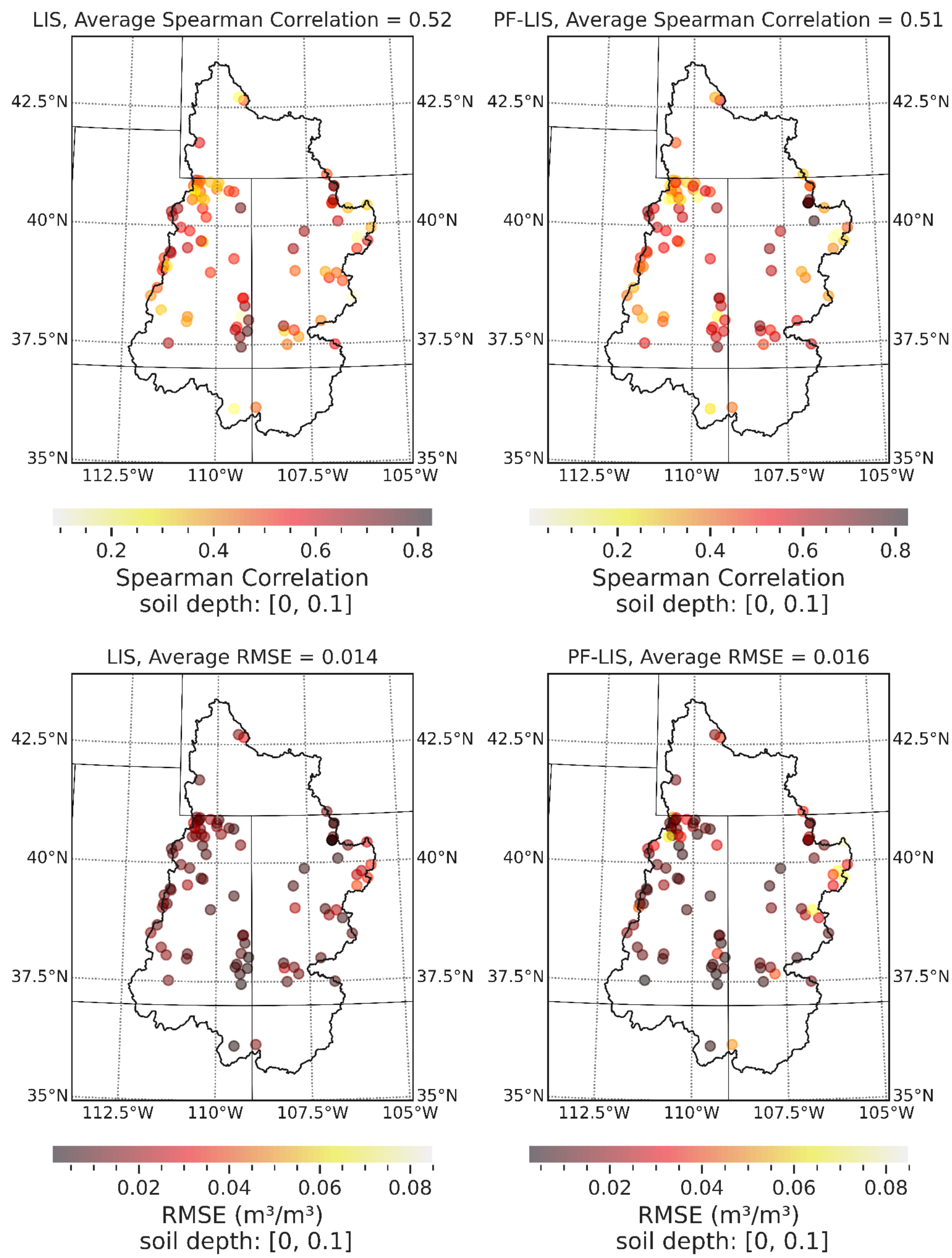

Figure 8. The Spearman's correlation coefficient and RMSE between the simulated and observed soil moisture at the soil depth of 0-0.1 m. This result is reported based on 20-year model simulation and observation data, from January 2002 to December 2022. At each station, performance metrics were calculated using 20 years of observations and model simulations. These metrics were then averaged across all stations to obtain a regional mean performance measure, which is reported in each plot.

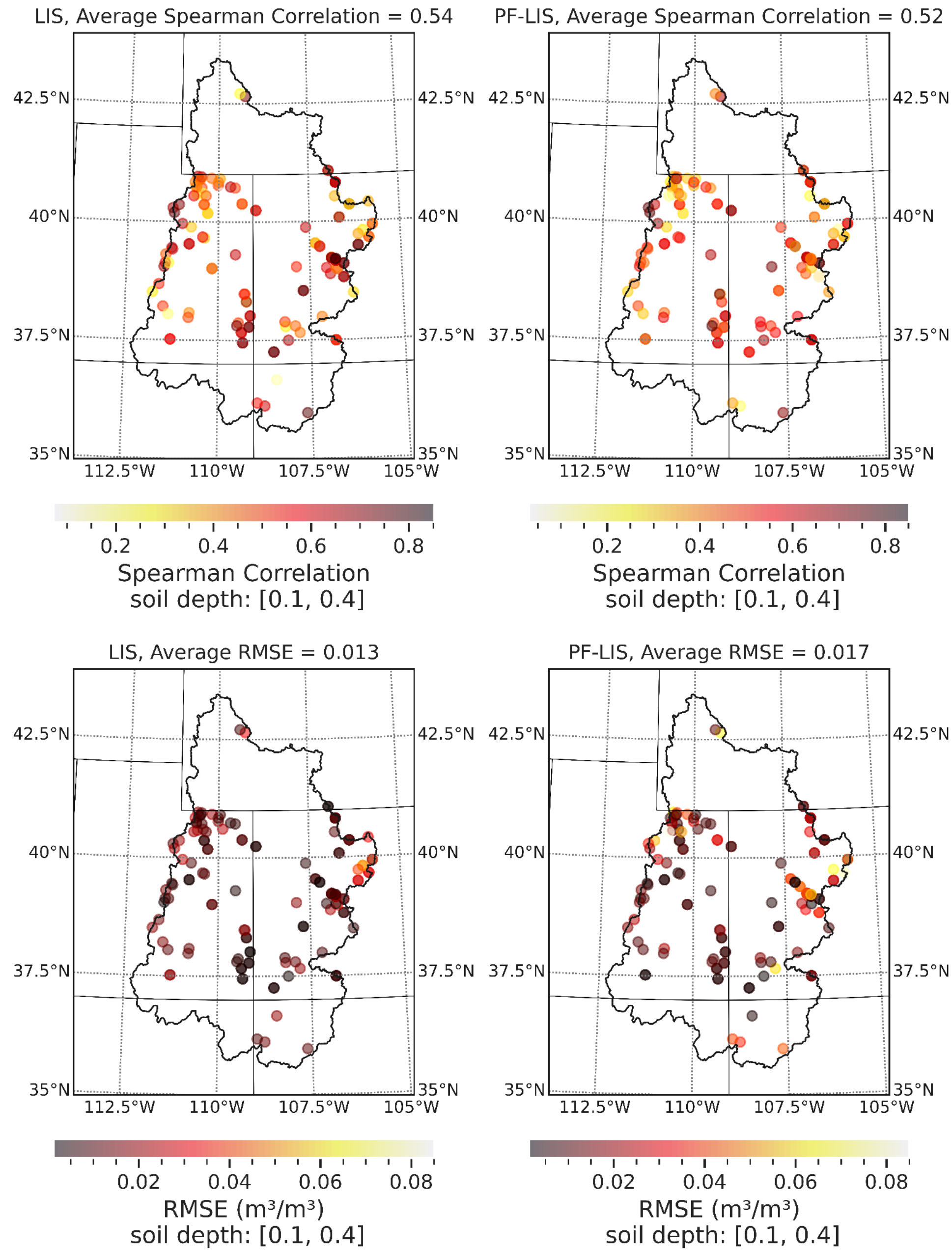

Figure 9. The Spearman's correlation coefficient and RMSE between the simulated and observed soil moisture at the soil depth of 0.1-0.4 m. This result is reported based on 20-year model simulation and observation data, from January 2002 to December 2022. At each station, performance metrics were calculated using 20 years of observations and model simulations. These metrics were then averaged across all stations to obtain a regional mean performance measure, which is reported in each plot.

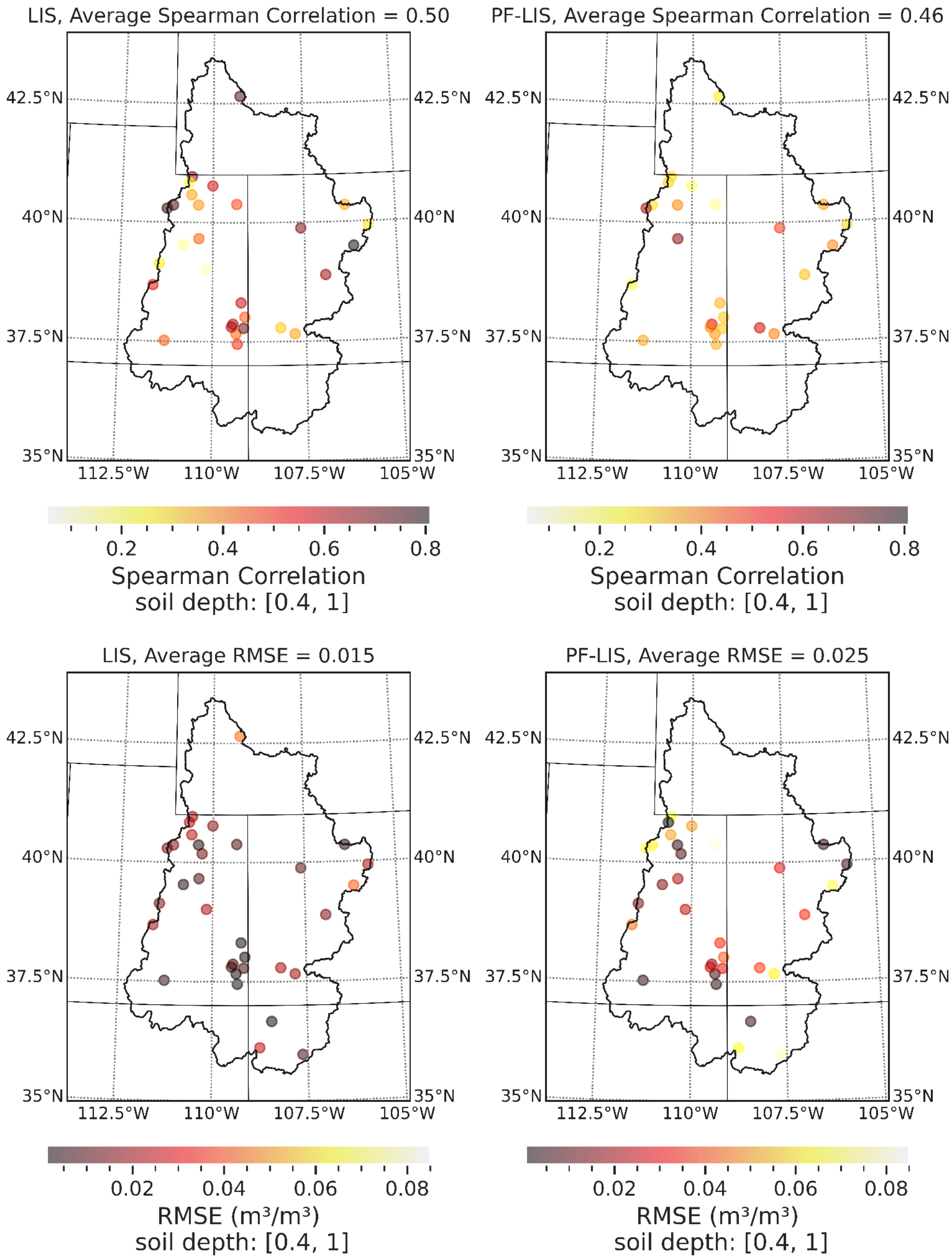

Figure S2. The Spearman's correlation coefficient and RMSE between the simulated and observed soil moisture at the soil depth of 0.4-1 m. This result is reported based on 20-year model simulation and observation data, from January 2002 to December 2022. At each station, performance metrics were calculated using 20 years of observations and model simulations. These metrics were then averaged across all stations to obtain a regional mean performance measure, which is reported in each plot.

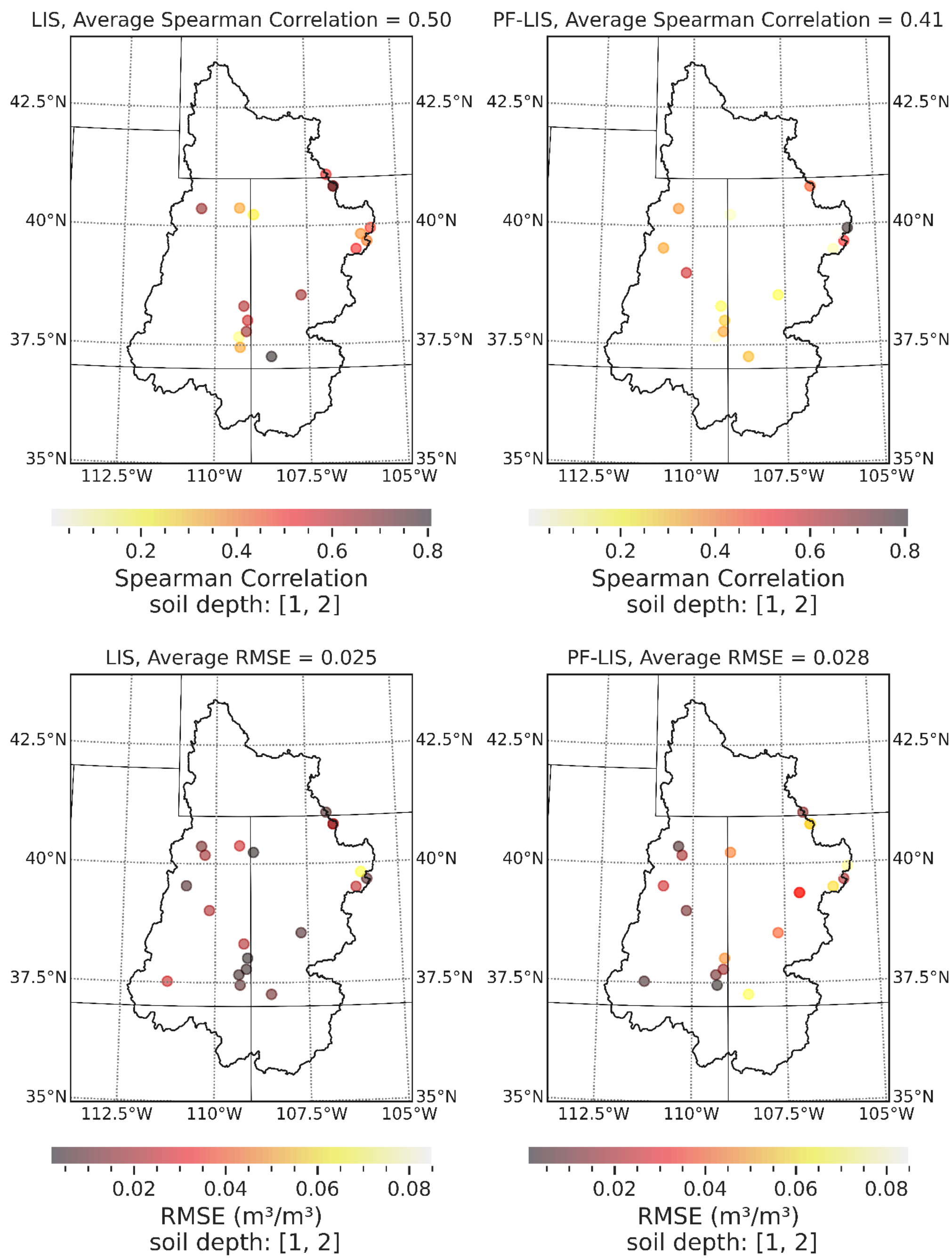

Figure S2. The Spearman's correlation coefficient and RMSE between the simulated and observed soil moisture at the soil depth of 1-2 m. This result is reported based on 20-year model simulation and observation data, from January 2002 to December 2022. At each station, performance metrics were calculated using 20 years of observations and model simulations. These metrics were then averaged across all stations to obtain a regional mean performance measure, which is reported in each plot.

Line 414, 415: So the simulations and observations are the mean soil moisture over the entire simulation period? The caption should specify precisely what data are shown.

At each station, performance metrics were calculated using 20 years of observations and model simulations. These metrics were then averaged across all stations to obtain a regional mean performance measure, which is reported in each plot.

To further address this comment, we added the above text to Figure 6 caption in the revised text.

"At each station, performance metrics were calculated using 20 years of observations and model simulations. These metrics were then averaged across all stations to obtain a regional mean performance measure, which is reported in each plot."

Line 418: It appears that in both cases, the statistical performance decreased for both metrics with the complex model. Can you say something about this? Are these differences meaningful, or is the slight drop "outweighed" by the improved spatial detail? Should the improved spatial detail not be expected to also give better fits, given the highly localized nature of the soil moisture observations?

The decrease in statistical performance (e.g., correlation) with ParFlow, despite its improved spatial representation, can be attributed to several factors: ParFlow explicitly resolves three-dimensional hydrological processes and captures fine-scale land surface characteristics better than Noah-MP. However, increased model complexity does not always translate to better statistical agreement with point-based observations, especially when observations are sparse or have localized variability that the model does not perfectly resolve. Soil moisture observations are highly localized, often influenced by microscale factors (e.g., vegetation, microtopography) that even high-resolution models like ParFlow may not capture accurately. In contrast, Noah-MP's parameterizations may act as an effective large-scale approximation, leading to a slightly better correlation with station-based observations. While Noah-MP shows slightly better accuracy in terms of correlation, the enhanced spatial realism in ParFlow is valuable for applications that require detailed hydrological representations. The slight statistical drop may not be as critical as the ability of ParFlow to resolve soil moisture variations across heterogeneous landscapes, making it more suitable for studies focusing on spatial processes rather than purely station-based validation.

To further address this comment, we added the above text to the revised manuscript.

“The decrease in statistical performance (e.g., correlation) with ParFlow, despite its improved spatial representation, can be attributed to several factors: ParFlow explicitly resolves three-dimensional hydrological processes and captures fine-scale land surface characteristics better than Noah-MP. However, increased model complexity does not always translate to better statistical agreement with point-based observations, especially when observations are sparse or have localized variability that the model does not perfectly resolve. Soil moisture observations are highly localized, often influenced by microscale factors (e.g., vegetation, microtopography) that even high-resolution models like ParFlow may not capture accurately. In contrast, Noah-MP’s parameterizations may act as an effective large-scale approximation, leading to a slightly better correlation with station-based observations. While Noah-MP shows slightly better accuracy in terms of correlation, the enhanced spatial realism in ParFlow is valuable for applications that require detailed hydrological representations. The slight statistical drop may not be as critical as the ability of ParFlow to resolve soil moisture variations across heterogeneous landscapes, making it more suitable for studies focusing on spatial processes rather than purely station-based validation.”

Line 447: Can contain …I guess your point is that these features can have an outsized influence? (They are, of course, also present in larger catchments too.)

We have revised the text as you suggested.

"While small drainage basins may experience water withdrawals and irrigation ditches, their smaller scale makes them more susceptible to localized anthropogenic influences, which can have an outsized impact on hydrological processes. In contrast, larger basins tend to buffer these localized effects, especially when considering monthly or annual timescales.”

Line 479: Subsurface!

Corrected. Thank you.

Line 483: This is not the best term. "Marginal" means minor or not important, or similar to something else. Maybe "limited" would be a better choice.

Revised. Thank you.

“…locations the model performance is limited due to the complex…”

Line 484: Could it not also be the complex geology, which the geological model and its numerical implementation cannot fully resolve? I.e., why put everything on the topography? If a considerable proportion of the mismatch is linked to the topography, then is this because the coarse grid (1 km) cannot capture it? If so, this should be stated explicitly. I guess it is partly this, but also the location mismatch (actual location vs. model centroid) plays a large part. This would be an argument in favor of employing an unstructured mesh, in which nodes can be placed to represent observation points exactly. (e.g., https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2020WR029390).

We agree with the reviewer and have revised the text to incorporate their suggestions and comments.

"Our comparison of water table depth estimates from the PF-LIS/Noah-MP model with those observed in USGS wells (refer to Table 1) reveals a general agreement between model simulations and observations. However, in some locations, model performance is marginal due to both the complex topography and geology of the UCRB. Additionally, the use of a coarse (1 km) grid may limit the model's ability to capture fine-scale topographic variations, and discrepancies between observation locations and model grid centroids further contribute to the mismatch."

Line 494: See above. Suggest using a different term.

Corrected. Changed to "limited".

Line 496, 497: Is this also because the geology is generally more complex in rugged areas, or mostly attributable to the fact that the spatial mismatch impacts are more pronounced there?

Yes, geology, of course, plays a significant role here. We have revised the text accordingly.

"Generally, the model's performance is contingent upon the geographical locations of the stations. Stations located in topographically complex surroundings, which often coincide with more complex geology, tend to yield lower model performance compared to those in flatter environments."

Line 508: Should mention that you are referring to spatially uniform here.

Revised. Thank you.

"...tend to have a more spatially uniform water table..."

Line 527, 528: Interesting observation!

Thanks for the feedback.